Research Product 87-25

U.S. Army Research Institute Officer Longitudinal Research Data Base (OLRDB) User's Manual

Leadership and Management Technical Area
Manpower and Personnel Research Laboratory

September 1987

U. S. Army Research Institute for the Behavioral and Social Sciences

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

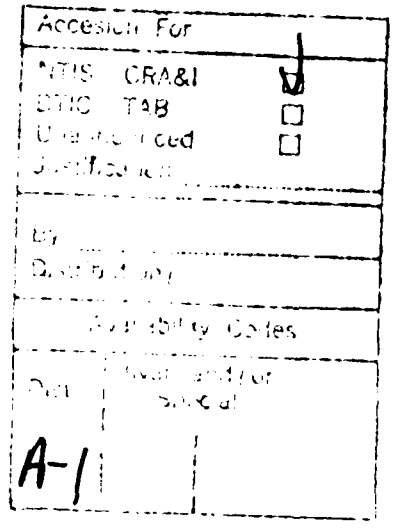
A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Technical review by

Robert Holz
David Horne
Trueman R. Tremble, Jr.



NOTICES

FINAL DISPOSITION: This Research Product may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: This Research Product is not to be construed as an official Department of the Army document in its present form.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Product 87-25	2. GOVT ACCESSION NO. ADA190588	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* U.S. ARMY RESEARCH INSTITUTE OFFICER LONGITUDINAL RESEARCH DATA BASE (OLRDB) USER'S MANUAL		5. TYPE OF REPORT & PERIOD COVERED Interim Report September 1986-September 1987 6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Fumiyo T. Hunter, Douglas L. Rachford, and Dennis R. Kelly (U.S. Army Research Institute) and Douglas Duncan (U.S.A. Information Systems Command--ARI)		8. CONTRACT OR GRANT NUMBER(s) --
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263743A794 452
11. CONTROLLING OFFICE NAME AND ADDRESS --		12. REPORT DATE September 1987 13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* --		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

--

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Longitudinal data base	User's manual
Officer training	Data dictionary
Officer utilization	Longitudinal research

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The U.S. Army Research Institute (ARI) Officer Longitudinal Research Data Base (OLRDB) supports systematic research on officer training, professional development, and utilization. The data base will be updated regularly to enable research on the current Army officer population, as well as analyses requiring historical tracking of information. The online OLRDB User's Manual and Data Dictionary were developed to facilitate access to the OLRDB data and to allow for an efficient updating of the Data Dictionary. This "hard copy"

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

ARI Research Product 87-25

20. Abstract (Continued)

of the manual serves as a general introduction to the data base. It contains the same material as the online manual, less the data dictionary. It is intended to help researchers, especially those without ready access to the ARI computer, to assess the potential usefulness of the OLRDB for their research.

Research Product 87-25

U.S. Army Research Institute Officer Longitudinal Research Data Base (OLRDB) User's Manual

Fumiyo T. Hunter, Douglas L. Rachford, and Dennis R. Kelly
Army Research Institute

Douglas Duncan
Scientific and Engineering Systems Division
U.S.A. Information Systems Command--ARI

Leadership and Management Technical Area
Robert F. Holz, Chief

Manpower and Personnel Research Laboratory
Newell K. Eaton, Director

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

September 1987

Army Project Number
2Q263743A794

Education and Training

FOREWORD

The Leadership and Management Technical Area of the U.S. Army Research Institute (ARI) conducts programmatic research to improve leader effectiveness, with a focus on the sequential, progressive development of leaders. To support this and other research, ARI is developing an Officer Longitudinal Research Data Base (OLRDB) along with an online user's manual and data dictionary stored in the ARI VAX computer. The data base will enable researchers to produce data-based information on officer training, professional development, and utilization.

The online OLRDB manual was developed to facilitate access to the data base and to provide an efficient means of incorporating changes resulting from regular updating of the data base. This "hard copy" of the user's manual has been prepared to serve as a general introduction to the data base. This version contains the same material as the online manual, less the data dictionary. It is intended to help researchers, especially those without ready access to the ARI VAX, to assess the potential usefulness of the data base for their research.

The development of the OLRDB has been briefed to the research sponsor, the Center for Army Leadership (29 April 1987), which recognizes its role as a research tool to generate information necessary for systematic enhancement of leader training and effectiveness.

EDGAR M. JOHNSON
Technical Director

U.S. ARMY RESEARCH INSTITUTE OFFICER LONGITUDINAL RESEARCH DATA BASE (OLRDB) USER'S MANUAL

CONTENTS

Preface

IMPORTANT NOTES

about the "Hard Copy" of the
U.S. Army Research Institute
Officer Longitudinal Research Data Base (OLRDB)
User's Manual

The U.S. Army Research Institute (ARI) Officer Longitudinal Research Data Base (OLRDB) will be updated on a regular basis to enable research on the current Army officer population, as well as analyses requiring historical tracking of information. The online User's Manual was developed to facilitate access to the OLRDB data. It also provides efficient means to update the Data Dictionary with new information.

This "hard copy" of the User's Manual has been prepared to serve as a general introduction to the data base, primarily for researchers without ready access to the ARI VAX. It is intended to help them assess the potential usefulness of the data base for their work. This version contains the same material as the online manual, less the Data Dictionary. Researchers who wish to examine the Data Dictionary need to do so by accessing the online manual in the ARI VAX computer.

The sections in the "hard copy" manual are not organized in the standard report format. They do, however, represent an exact copy of the online manual, intended to assure comparability for users referring to the online and the "hard copy" versions.

The readers of the "hard copy" are reminded to note the date of this preface and inquire about an updated version, if the date appears old.

Fumiyo T. Hunter
OLRDB Project Manager

AV 284-8293
(202) 274-8293

September 1987

U.S. ARMY RESEARCH INSTITUTE

OFFICER LONGITUDINAL RESEARCH DATA BASE (OLRDB)

USER'S MANUAL

INTRODUCTION

The Online User's Manual for the Officer Longitudinal Research Data Base (OLRDB) was designed to facilitate ARI researchers' access to data in the OLRDB. The manual consists of two parts. The first is the text file that you are now reading that describes:

(1) the data base
(purpose, contents, organization, how to request data, references),
(2) how to use the Data Dictionary, and
(3) how to run a job using data from the data base.

The second part is the Data Dictionary. The dictionary provides information about each data element, or variable, in the OLRDB, including which data set contains the variable, description of the variable, how it is coded, and the descriptions of the codes.

To use the Data Dictionary and the OLRDB, a researcher should first become familiar with the information in the User's Manual. Then, by looking through the Data Dictionary (see "How to Use the Data Dictionary" section of this Manual), one can determine the availability of desired information in the OLRDB. Once the researcher selects the OLRDB data elements to be analyzed, the Data Request Form, shown later in this Manual, will be submitted to the Data Base Manager. The examples in the "How to Run a Job" section will assist the researcher in compiling the computer commands needed to construct a working file and analyze data using the Statistical Analysis System (SAS) at the National Institutes of Health (NIH) computer facility. Researchers using the NIH computer would also be able to apply other statistical packages such as SPSS and BMDP to SAS-format working files.

Knowledge of statistical packages, particularly SAS, is helpful. Understanding how to use the NIH computer is also helpful. Full instruction in these areas is beyond the scope of this User's Manual, but excellent documentation is available on both SAS and the NIH-IBM system. Courses are also available on these topics. If you need additional information or have any comments about the accuracy, readability, or helpfulness of the User's Manual,

contact the OLRDB Manager--Fumiyo Hunter (202-274-8293, AV 284-8293, ARI VAX Username = OLRDB).

DESCRIPTION OF THE OLRDB

Purpose

The purpose of the Officer Longitudinal Research Data Base (OLRDB) is to make personnel, pre- and post-commissioning training, and field performance data on U.S. Army officers available for research. The OLRDB was constructed to provide efficient access to these data with the best possible documentation. Historical data from various Army agencies have been incorporated into several SAS data sets, organized by similarity of contents, with variable and code labels. This Online User's Manual and a comprehensive Data Dictionary are provided to assist researchers in using the data base most effectively.

Contents

1. Officer Personnel Data. The OLRDB began by converting yearly "snapshots" (1979-1985) of the Officer Master File (OMF) compiled by U.S. Army Military Personnel Center into SAS data sets and describing these data in the Data Dictionary.

Personnel data from the Defense Manpower Data Center (DMDC) Master and Loss File, containing separation information for the period of 1970-1985, were used to cross check the accuracy of data from the 1979-1985 OMF and to obtain OMF-equivalent data for 1970-1978. The OLRDB Core Data Set was then constructed to (1) provide an accurate, comprehensive list of commissioned officers on active duty during the period of 1970-1985, (2) provide one data set with the most commonly used data, and (3) create an encrypted list of identifying information to make it possible to link individual data from various data sets while maintaining a level of privacy.

All OLRDB data sets are updated yearly. Thus, all commissioned officers on active duty are covered for the period 1970 to present. The majority of cases are in the Regular Army or Active Duty Reserve forces.

2. Pre-Commissioning Data. Data from precommissiong sources such as Reserve Officers' Training Corps (ROTC), Officer Candidate School (OCS), and the U.S. Military Academy (USMA) are being added as they are obtained. ROTC Advanced Camp performance and Commissioning File data collected during 1982-1985 have been included into the OLRDB. The USMA and OCS data sets will be established in 1987-1988.

3. Post-Commissioning Data. Information from the Automated Instructional Management System (AIMS), a TRADOC data management system containing Officer Basic and Advanced Course grades, will be incorporated during 1987-1988. Leadership effectiveness and unit performance data from field training exercises and garrison settings will be added as they become available.

Organization of Data

The OLRDB data reside on a set of magnetic tapes at the National Institutes of Health computer facility with an IBM mainframe, while the User's Manual and Data Dictionary are on-line on the ARI VAX computer.

The OLRDB can be thought of as one big data set divided into pieces. A simple view might look like this:

	Variables				
Individuals	Sex	Rank	Branch	Marital Status	Date of Commission
1	M	CPT	AR	M	14 JUN 74
2	F	1LT	AR	S	Ø3 AUG 71
3	M	MAJ	IN	D	27 JAN 71
4	M	LTC	IN	S	18 DEC 68
5	F	MAJ	MI	M	Ø9 NOV 73
6	M	2LT	FA	D	23 JUN 85

Each individual has one record containing all available data for that individual. However, because of the large number of variables (>1K), the variables have been divided into 18 data sets which can be merged together for each individual, as of September 1987. All but one of these data sets are in SAS format, with variables and data values already labeled.

1. Personnel Data Sets. Data from the OMF and Master and Loss File are divided into 14 groups (such as biographical data, assignment history, promotion history, military and civilian education, branch/functional area/skills, and awards) and stored on 14 magnetic tapes. Each of these tapes contains the most recent information on all active duty records from 1979 to present. For officers who have separated during this period, the information at the time of separation is retained. For officers presently on active duty, the most current information, updated yearly, is retained in the SAS data sets.

The 14 data sets described above collectively contain most of the variables in the OMF. In contrast, the two core data sets include selected OMF variables frequently used for research and

contain the most rigorously verified listing of active duty cases from 1970-present. The first of these is the SAS Core Data Set. Like the other 14 SAS data sets, the SAS Core Data Set contains the most recent information available for each record. As personnel information on individuals changes over the years, the new information replaces the old in these data sets. (See ARI Research Product, entitled Development of Core Data Set of the Officer Longitutidnal Research Data Base, 1987, by D. Younkman, Fu Associates, Ltd.)

The second core data set is the Longitudinal Data Set, the only OLRDB data set not in the SAS format. This data set contains the same variables as the SAS Core Data Set, but in a "cleaned" up, raw form. In addition, this data set repeats the same set of variables (with different values when they change) for each year from 1970 to present. This feature permits tracking of changes in the values of any given variable over the years and the approximate timing of the changes. It also allows application of computer programs other than SAS.

LONGITUDINAL DATA SET (Contains yearly core data set variables)

1970 DATA	1971 DATA	1972 DATA	1973 DATA	1974 DATA	1975 DATA	. . .

2. Pre-Commissioning Training Data. There are 2 ROTC data sets in SAS format, one containing the ROTC Advanced Camp performance records and the other, the final program records (Commissioning File information) from 1982 to present. (See ARI Research Product, entitled Development of ROTC Data Sets and Evaluation of Their Usefulness for Officer Longitudinal Research Data Base, 1987, by D. Younkman, Fu Associates, Ltd.) Unlike the personnel data sets in which the same records are updated yearly, the training data sets contain different sets of individuals in each yearly segment. For example, the Advanced Camp data for Officer X would be found in the Advanced Camp data set for the year that he/she attended the Advanced Camp.

The academic records from the U.S. Military Academy (USMA) and Officer Candidate School (OCS) will be structured similarly to the ROTC data sets and integrated into the OLRDB in 1987-1988.

3. Post-Commissioning Training Data. As of June 1987, work was underway to collect the Officer Basic and Advanced Course (OBC, OAC) performance data from the Automated Instructional Management System, Training and Doctrine Command, and to integrate them into the OLRDB. When completed, these data will be organized by branch school, course (OBC and OAC), and year. As with the pre-commissioning data sets, an individual's performance in OBC or

OAC will be entered once, for the year the course is taken, and not revised yearly. However, some individuals will be included in the OBC data set and, later, in the OAC data set.

4. Data Dictionary. There is one online Data Dictionary describing data elements from all OLRDB data sets, with specifications as to which data set and tape contain each data element. One can query the Data Dictionary to list names of all data sets, all variables in a given data set, all data values for a given data element, etc., and print out any part of these lists as hard copies. Additional considerations might be: In what years was particular information collected? Are there known problems with the variables or codes? Are there related variables a researcher might want to consider? The Data Dictionary contains answers to many of these questions. See instructions for entering the Data Dictionary at the end of this User's Manual.

If researchers know what variables they are looking for, they can enter the Data Dictionary, examine the descriptions and codes for those variables, and determine the usability of the data and what data set(s) are needed. Queries on keywords or codes are possible if the variables are not known in advance. One can also scan the data sets, variables, and codes to get an idea of what data are included. Instructions for querying the Data Dictionary and examining variables are provided once you have entered the Data Dictionary.

Accessing and Merging Data Sets

Researchers with NIH accounts may access the OLRDB data sets at the NIH computer facility. The OLRDB data tapes can be read but are protected from further manipulation. The standard procedure for users is: (1) to extract from the data sets the variables selected for a particular research need, (2) to create a working file of the extracted data to be stored under the researchers' account, and (3) to scratch the working file when the necessary analyses are completed, unless yearly update and continued use of the working file is planned. Researchers are responsible for creating the working file and using it in accordance with the requirements that are specified in the Data Request Form shown later.

To protect the privacy of individuals, social security numbers have been encrypted and other personal identifiers have been stripped from the OLRDB data sets, and a linking file of encrypted information has been constructed. Each record in each data set contains an encrypted personal identifier (MATCHCOD). Variables from several OLRDB data sets can be merged together by the matchcode to create a working data set with just the data needed for a particular research project.

A most powerful aspect of the OLRDB is that officer data from other research projects can be merged with the OLRDB, greatly expanding the scope of possible analysis and multiplying the usefulness of research efforts. However, individual social security numbers would be needed to merge OLRDB data sets with non-OLRDB data. Requests for converting the OLRDB Matchcodes to social security numbers should be addressed to the OLRDB Manager via OLRDB Data Request Form discussed later.

Important Reminders

It is incumbent on the researcher to carefully screen the Data Dictionary to understand clearly what variables are available and what each means. Not all data are available on each individual. Some variables were not collected for every year; coding of some variables changed over the years; and historical records provide less information the further back one looks.

Currently, inclusion of records in the OLRDB is limited to commissioned officers in active duty, including those in the Regular Army, Active Duty Reserve, and Army of the U.S. Non-commissioned and warrant officers, as well as reserve officers not in active duty, are excluded.

If users have questions about the OLRDB or the documentation, they are encouraged to contact the OLRDB Manager, currently Fumiyo Hunter, 202-274-8293, AV 284-8293, or send e-mail messages to ARI VAX Username = OLRDB.

How to Request OLRDB Data

After a researcher selects the OLRDB variables to use, he/she must complete the following Data Request Form and submit it to the Data Base Manager, Fumiyo Hunter, Army Research Institute, Leadership and Management Technical Area, PERI-RL, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600. The form will serve two purposes: (1) to keep records of OLRDB use (who uses it, frequently used types of data, etc.) and (2) to assure adherence to the requirements of the Privacy Act. The entire Data Request Form is presented below. A researcher may print out the form from the User's Manual or request a hard copy from the Data Base Manager.

REQUEST TO ACCESS/TRANSFER DATA

FROM OFFICER LONGITUDINAL RESEARCH DATA BASE (OLRDB)

GENERAL INFORMATION

1. Use of data contained in the Officer Longitudinal Research Data Base (OLRDB) is hereby requested.

2. I have read and understood the information contained in the on-line OLRDB User's Manual provided on the Army Research Institute VAX computer.

3. It is understood that OLRDB contains the type of personal information on U.S. Army officers covered by the 1974 Privacy Act and is intended for research use primarily by DoD researchers. Researchers using the data base are responsible for preventing disclosure of any specific information about any particular individual to anyone, including the individual him/herself, Army agencies, and any other person or organization.

4. It is understood that the user will construct a working file of selected cases and data elements from the OLRDB and store the working file under his/her account. Exceptions will be made for users with no NIH account, from distant locations. Such users may request a transfer of data to other computer facilities by magnetic tape. The OLRDB data sets are to be "Read-Only" and to serve as the source of working files.

5. Any publication resulting from research using the OLRDB data will include an acknowledgement of the OLRDB, and a copy of such publication will be provided to the OLRDB Manager.

DESCRIPTION OF RESEARCH PROJECT

6. Nature of research for which OLRDB data are being requested:

_____

_____

_____

7. The OLRDB data requested will be merged with data generated from other projects/sources.

_____ No

_____ Yes Describe the nature of data to be combined with the OLRDB data _____

_____

_____

_____

SPECIFICATION OF DATA REQUESTED

8. The OLRDB data will be accessed by:

a. Direct reading of the SAS data sets stored at the NIH computer facility

_____ No. If no, go to 8.b.

_____ Yes. If yes, complete 8.a.1. below.

8.a.1. The working file will be downloaded to computer system other than the NIH system.

_____ No, it will remain at NIH

_____ Yes, it will be downloaded to:

_____
Name/location of the receiving system

b. Request data extracts to be transferred to another computer facility by magnetic tape(s). The tape with OLRDB data extracts should be mailed to:

Name: _____

Address: _____

_____

_____

Phone Number: _____

Tape specifications:

Track/BPI: _____

_____ EBCDIC _____ ASCII

_____ Labeled _____ Non-labeled

Other: _____

_____

9. The data elements to be extracted into a working file or onto a tape are listed on the Attachment.

10. The OLRDB records need to be identified with social security numbers to merge them with data from other sources.

_____ Yes. If yes, the Data Base Manager will convert the encrypted Matchcode for requested OLRDB cases and data sets into social security numbers.

_____ No

11. If the request calls for a working file with the encrypted Matchcode converted into social security numbers to merge with non-OLRDB data, it is understood that once social security numbers are used to merge data, the following steps will be taken to reduce possibilities of accidential disclosure of personal information:

a. Each record will be assigned a unique sequence number.

b. An individual's social security number will be associated with the sequence number.

c. A separate file containing the matched social security and sequence numbers will be created and stored in a separate file with appropriate security measures.

d. Personal identification information used for merging data sets, including social security numbers, will be stripped from the working file of the merged data.

12. Working files created at NIH, ARI, or other computer facilities and those residing on tapes will be:

_____ Scratched, when the current use is completed.

_____ Saved for future use. Explain the nature of the future use: _____

_____

_____

13. Future updates of the longitudinal data are requested.

_____ Yes _____ No _____ TBD

Submitted by: _____
Name, title

Signature: _____

Organization: _____

Date: _____

Submitted to: Fumiyo Hunter, OLRDB Manager
Army Research Institute
PERI-RL
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

(202) 274-8293, AV 284-8293

ATTACHMENT

A. Sample specification (e.g., all cases in the OLRDB, active duty since 1980, ROTC graduates only, etc.):

B. OLRDB Data Elements To Be Extracted

	SAS DATA SET NAME	VARIABLE NAME
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		
15.		

Continue on additional sheets.

References

A. The following reports contain detailed information on the development, structure, and contents of respective OLRDB data sets.

Rachford, D. L. (1984). Building the Officer Longitudinal Data Base (OLRDB) (ARI/LMTA Working Paper 84-07).

Younkman, D. D. (1987). Development of Core Data Set of The Officer Longitudinal Research Data Base (ARI Research Product 87-22). Alexandria, VA: U.S. Army Research Institute. (In preparation)

Younkman, D. D. (1987). Development of ROTC Data Sets and Evaluation of Their Usefulness for Officer Longitudinal Research Data Base (ARI Research Product 87-24). Alexandria, VA: U.S. Army Research Institute. (In preparation)

Hunter, F. T. (1987). Officer Longitudinal Research Data Base: Development and Utilization. Paper presented at the 29th Military Testing Association Conference. Ottawa, Canada.

B. The following reports describe research/data analysis which utilized OLRDB data.

Hunter, F. T. (1987). Relationships Among Standardized Literacy Test Scores, ROTC Grades, and Officer Basic Course Performance (ARI/LMTA Working Paper 87-01).

Kelly, D. R. (1986). A Comparison of the OLRDB and Defense(yr) Almanacs on Marital Status of Army Officers (ARI/LMTA Working Paper 86-02).

Rachford, D. L. (1985). Gender Analysis of the Professional Development of Officers Study (PDOS) Survey (ARI Research Report 1409). Alexandria, VA: U.S. Army Research Institute. (AD A168 878)

Rachford, D. L., Twohig, P., & Rigby, C. K. (1984). Tabulations on Selected Career Variables from the Prototype ARI Officer Longitudinal Data Base (OLDB) (ARI/LMTA Working Paper 84-05).

Acknowledgment

Many individuals are reponsible for the development of the OLRDB. Strong support and foresight by the Manpower and Personnel Research Laboratory (MPRL) Directors (Drs. Joyce Shields and Kent Eaton) and the Leadership and Management Technical Area Chief (Drs. William Haythorn and Robert Holz) were instrumental in its development. Doug Rachford, Dennis Kelly, and Fumiyo Hunter were mainly responsible for the overall design, data gathering, and construction of the OLRDB. Doug Ducan and Neal Maclay provided the programming for the User's Manual and Data Dictionary. Substantial efforts were also made by Dana Kruser, Susan Smits, and Francis Grafton. Army agencies such as MILPERCEN, DMDC, ROTC, TRADOC, and U.S. Military Academy have cooperated and provided much asistance in developing the OLRDB. Lastly, the staff of Fu Associates, Ltd. contributed consistently high quality work on the development and maintenance of the OLRDB.

HOW TO USE THE DATA DICTIONARY

The Data Dictionary is the key to using the OLRDB. The Data Dictionary allows researchers to examine what data are available and whether they will be useful for specific research. To use the Data Dictionary, users should return to the online User's Manual main menu. When the data-dictionary option is selected from the main menu, a list of all OLRDB data sets appears on the screen. The Data Dictionary menu at the bottom of the screen instructs the users on how to select a data set and proceed to the list of variables, variable descriptions, codes, and code labels. It also contains options for examining other data sets or returning to the User's Manual main menu. The online instructions and menu should assist users to review the OLRDB data thoroughly. However, users are encouraged to contact the OLRDB Manager with any questions or suggestions.

HOW TO RUN A JOB

Running a job using data from the Officer Longitudinal Research Data Base (OLRDB) will require a basic knowledge of Statistical Analysis System (SAS), preferably at NIH. This section will give guidelines and examples for running some typical jobs using variables taken from the OLRDB data sets. This section is not an attempt to teach researchers how to use SAS at NIH or on the VAX system. SAS consultants at NIH (202-496-6037) provide assistance with SAS-related questions. NIH also has a PAL unit (202-496-5525) that will help with Job Control Language (JCL) or problems in running jobs. Courses in SAS are offered at NIH or through the SAS Institute.

Running a job at NIH will require the following pieces of information which can be acquired from the Data Dictionary.

1. Variable names.
2. The name of the OLRDB SAS data set which contains the variables to be used.
3. The tape name and number associated with the SAS data set.

The OLRDB data are stored at NIH on tapes. Each tape has a six digit tape number and a name, referred to in the JCL as a DSN. For storage reasons and simplicity of design, each OLRDB tape contains one OLRDB data set with the respective variables associated with that data set. Remember that the "tape name" and "SAS data set name" may be different; the tape name is needed for the JCL, the data set name, for use with SAS commands.

The information listed above needs to be "plugged into" the JCL before running a job at NIH. From the Data Dictionary, the user may write down some information while using the Dictionary or print out the lists of variables, data set names, tape names, and tape numbers. Instructions for printing are available in the Dictionary.

The general procedure for using the OLRDB data will be:

(1) extract selected variables from appropriate data sets,
(2) merge data from two or more data sets, if needed,
(3) create a working file, registered under the user's account, and
(4) store the working file until the completion of analytic work **AT WHICH TIME IT MUST BE SCRATCHED**, unless continued use is indicated in the Data Request Form.

Some OLRDB data sets contain large numbers of cases, spanning many years. Many of them may not be applicable for a particular analysis purpose. Selecting only the necessary cases for the working file before proceeding with data analysis will help to reduce computer time and cost.

Example Job Set-ups

The following examples are basic JCL structures that show some of the various ways of accessing, storing, and using the OLRDB data.

Note: Several JCL commands are necessary for jobs using the OLRDB data tapes. They are:

```
/*ROUTE XEQ TAPE
/*MESSAGE tapenumber,R
/*ACCESS WRZ1KFD
//SASLIB   DD DSN=WRZ1KFD.OLRDB.FORMATS,DISP=SHR
```

/*ROUTE XEQ TAPE is needed for all jobs requiring tape(s) to be mounted. After /*MESSAGE, list the tape number and a code "R" indicating the tape is to be read but not written. The OLRDB tapes are specified to be "read-only" to prevent accidental alteration of the data. After /*ACCESS, list the NIH account/initial WRZ1KFD under which the tapes are registered. //SASLIB line identifies the SAS format library (a computer file) that contains formats, or variable value descriptions, for many of the OLRDB variables. The variables are stored in the SAS data sets along with these formats. Therefore, attemps to access OLRDB SAS data sets without referring to the format library by this JCL statement will be aborted.

Example 1: Create a working file of selected variables on disk and analyze data. This example shows the JCL and procedure that could be used to access one tape, temporarily store the variables selected from that tape, and run frequencies on the variables. In this example, the variables named BZMAJ, BZLTC, BZCOL, and TGRA are being pulled from the data set named RAN85A (LINE 14). The variables are contained on the tape titled, WRZ1KFD.OMF.SDSD (line 7). The information on these variables is being temporarily stored, under the user's account, on TMP002 (lines 9/10) in the data set BZRANK (line 13). This temporary data set will be scratched from the disk in 2-3 days. The TMP disk is used only for short term (temporary) storage. After the variables have been pulled from the source tape (# 034221) and put onto a disk for use (in this example TMP002), the analyses that can be performed are almost unlimited. In Example 1, frequencies are computed.

```
1  //XXXEX1    JOB  (YYYY,860,B),KELLY
2  /*ACCESS WRZ1KFD
3  //PROCLIB     DD  DSN=ZABCRUN.PROCLIB,DISP=SHR
4  //STEP1       EXEC SAS,REGION=4000K
5  /*ROUTE       XEQ TAPE
6  /*MESSAGE     034221,R
7  //FYALL1 DD DSN=WRZ1KFD.OMF.SDSD,DISP=SHR,LABEL=(1,SL),
8  //  UNIT=TAPE,VOL=SER=(034221)
9  //TEMP DD DSN=YYYYXXX.BZRANK,UNIT=FILE,DISP=(NEW,KEEP),
10 //  VOL=SER=TMP002,SPACE=(TRK,(5,5),RLSE)
11 //SASLIB      DD DSN=WRZ1KFD.OLRDB.FORMATS,DISP=SHR
12 //SYSIN       DD  *
13 DATA  TEMP.BZRANK;
14 SET FYALL1.RAN85A(KEEP=BZMAJ BZLTC BZCOL TGRA);
15 PROC FREQ;TABLES
16       BZMAJ*TGRA
17       BZLTC*TGRA
18       BZCOL*TGRA
19       BZMAJ*BZLTC*TGRA
20       BZMAJ*BZCOL*TGRA
21       BZMAJ*BZLTC*BZCOL*TGRA;
```

Example 2: Create and store selected variables on disk as a permanent file. This example is identical to Example 1 with the exception of lines 9, 10 and 13. By storing the data on disk, the user will have a permanent file which can be accessed indefinitely. However, the data storage costs are the highest for the public disks. All but the smallest size working files should be stored on the MSS or magnetic tapes.

```
1  //XXXEX2    JOB  (YYYY,860,B),KELLY
2  /*ACCESS WRZ1KFD
3  //PROCLIB     DD  DSN=ZABCRUN.PROCLIB,DISP=SHR
4  //STEP1       EXEC SAS,REGION=4000K
5  /*ROUTE       XEQ TAPE
6  /*MESSAGE     004221,R
7  //FYALL1 DD DSN=WRZ1KFD.OMF.SDSD,DISP=SHR,LABEL=(1,SL),
8  //  UNIT=TAPE,VOL=SER=(004221)
9  //PERM DD DSN=YYYYXXX.BZRANK,UNIT=FILE,DISP=(NEW,CATLG),
10 //  VOL=SER=FILE44,SPACE=(TRK,(5,51),RLSE)
11 //SASLIB      DD DSN=WRZ1KFD.OLRDB.FORMATS,DISP=SHR
12 //SYSIN           DD   *
13  DATA    PERM.BZRANK;
14 SET FYALL1.RAN85A(KEEP=BZMAJ BZLTC BZCOL TGRA);
15 PROC FREQ;
16   TABLES
17      BZMAJ*TGRA
18      BZLTC*TGRA
19      BZCOL*TGRA
20      BZMAJ*BZLTC*TGRA
21      BZMAJ*BZCOL*TGRA
22      BZMAJ*BZLTC*BZCOL*TGRA;
```

Example 3: Select variables from two data sets and store merged working file on Mass Storage System (MSS). The example below allows the user to use variables from two data sets (lines 8-11) on tapes and store the output permanently on the MSS, under the user's account (lines 12-13). MSS is useful for storing large working data sets that will be accessed frequently. The user has to indicate in the JCL two tape numbers (lines 9, 11), and two tape names (lines 8, 10). Line 17 is a SAS statement indicating which variables and respective data sets are to be used. The KEEP statement will read and store the variables HEIGHT, on the data set named BIO85A, and TGRA on the data set named RAN85A. These two variables will be on the newly created data set named HGTRANK (line 16) on MSS (lines 12-13).

```
1  //XXXEX3    JOB  (YYYY,860,B),KELLY
2  /*ACCESS WRZ1KFD
3  /*ROUTE XEQ MSS
4  //PROCLIB     DD  DSN=ZABCRUN.PROCLIB,DISP=SHR
5  //STEP1       EXEC SAS,REGION=4000K
6  /*ROUTE   XEQ TAPE
```

```
7  /*MESSAGE Ø12739,R;Ø34221,R
8  //FYALL1  DD DSN=WRZ1KFD.OMF.SDSA,DISP=SHR,LABEL=(1,SL),
9  //  UNIT=TAPE,VOL=SER=(Ø12739)
10 //FYALL2  DD DSN=WRZ1KFD.OMF.SDSD,DISP=SHR,LABEL=(1,SL),
11 //  UNIT=TAPE,VOL=SER=(Ø34221)
12 //PERM    DD DSN=YYYYXXX.HGTRANK,UNIT=MSS,DISP=(NEW,CATLG),
13 //  SPACE=(CYL,(5,5),RLSE)
14 //SASLIB  DD DSN=WRZ1KFD.OLRDB.FORMATS,DISP=SHR
15 //SYSIN   DD  *
16 DATA PERM.HGTRANK;
17 MERGE FYALL1.BIO85A(KEEP=HEIGHT) FYALL2.RAN85A(KEEP=TGRA);
18 BY MATCHCOD;
19 PROC FREQ;
20   TABLES HEIGHT*TGRA;
```

Once the user has stored information on a temporary file, disk, or the MSS it can be used again with a much faster job turnaround time. However, long-term storage of large data sets on MSS or public disks is not recommended.

Example 4: Analyze data stored on MSS. The example below retrieves data from an existing data set named HGTRANK (line 6), stored on MSS in Example 3. Because the data set is already on the MSS, the user can access it quickly (no tapes are required), manipulate the data, and perform analyses without recreating the working file. Note that HGTRANK2 in line 9 is a temporary data set that will be used only for the duration of the job. A correlation between two of the variables in the data set on MSS is calculated after the alphabetical codes for Variable TGRA are converted to numbers.

```
1  //XXXEX3    JOB  (YYYY,860,B),KELLY
2  /*ACCESS WRZ1KFD
3  /*ROUTE XEQ MSS
4  //PROCLIB    DD  DSN=ZABCRUN.PROCLIB,DISP=SHR
5  //STEP1      EXEC SAS,REGION=4ØØØK
6  //FYALL1  DD DSN=YYYYXXX.HGTRANK,DISP=SHR,UNIT=MSS
7  //SASLIB  DD DSN=WRZ1KFD.OLRDB.FORMATS,DISP=SHR
8  //SYSIN   DD  *
9  DATA HGTRANK2;
10 SET FYALL1.HGTRANK;
11 IF TGRA='WO1' THEN TGRA1=1;
12 IF TGRA='CW2' THEN TGRA1=2;
13 IF TGRA='CW3' THEN TGRA1=3;
14 IF TGRA='CW4' THEN TGRA1=4;
15 IF TGRA='2LT' THEN TGRA1=5;
16 IF TGRA='1LT' THEN TGRA1=6;
17 IF TGRA='CPT' THEN TGRA1=7;
18 IF TGRA='MAJ' THEN TGRA1=8;
19 IF TGRA='LTC' THEN TGRA1=9;
20 IF TGRA='COL' THEN TGRA1=10;
```

```
21 IF TGRA='B G' THEN TGRA1=11;
22 IF TGRA='M G' THEN TGRA1=12;
23 IF TGRA='LTG' THEN TGRA1=13;
24 IF TGRA='GEN' THEN TGRA1=14;
25 IF TGRA='G A' THEN TGRA1=15;
26 PROC CORR VAR TGRA1;
27 WITH HEIGHT;
```

Example 5: Merge variables from a tape and the MSS by MATCHCOD, create a new data set to store on tape. This example takes the variable BABR off of the data set BRN85A and the variable TGRA from the data set HGTRANK (line 16). The first data set is contained on tape number 034050, named WRZ1KFD.OMF.SDSB (lines 8-9). The second data set has been stored as an MSS data set HGTRANK (line 10) in the Example 3. Recall that the data set HGTRANK contains data from two other data sets. By merging HGTRANK with BRN85A, a new separate data set has been created (BABRTGRA, line 15) with variables from three separate data sets and stored on tape (lines 11-12) under the user's account. NIH computer users are limited to using two tapes in a B class job. By merging previously stored data sets (derived from the data sets on tapes) with data sets on one or two tapes, they can avoid having to run C class jobs. C class jobs only run overnight. (See the NIH Computer User's Guide for descriptions of job classes.)

```
1  //XXXEX5    JOB  (YYYY,860,B),KELLY
2  /*ACCESS WRZ1KFD
3  /*ROUTE XEQ MSS
4  //PROCLIB    DD  DSN=ZABCRUN.PROCLIB,DISP=SHR
5  //STEP1      EXEC SAS,REGION=4000K
6  /*ROUTE   XEQ TAPE
7  /*MESSAGE 034050,R;999999,W
8  //FYALL1  DD DSN=WRZ1KFD.OMF.SDSB,DISP=SHR,LABEL=(1,SL),
9  //  UNIT=TAPE,VOL=SER=(034050)
10 //FYALL2  DD DSN=WRZ1KFD.HGTRANK,DISP=SHR,UNIT=MSS
11 //PERM    DD DSN=YYYYXXX.BABRTGRA,UNIT=TAPE,DISP=(NEW,KEEP),
12 //  VOL=(PRIVATE,SER=999999)
13 //SASLIB  DD DSN=WRZ1KFD.OLRDB.FORMATS,DISP=SHR
14 //SYSIN   DD  *
15 DATA  PERM.BABRTGRA;
16 MERGE FYALL1.BRN85A(KEEP=BABR) FYALL2.HGTRANK(KEEP=TGRA);
17 BY MATCHCOD;
18 PROC FREQ;
19   TABLES
20      BABR*TGRA;
```